Docket No. DIGI8

RAID SYSTEM HAVING CHANNEL CAPACITY UNAFFECTED BY ANY SINGLE COMPONENT FAILURE

CROSS-REFERENCE TO RELATED APPLICATIONS.

Not Applicable.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT.

Not Applicable.

Reference to a "Microfiche appendix."

Not Applicable.

BACKGROUND OF THE INVENTION.

FIELD OF THE INVENTION

[0001] This invention relates to RAID systems with provisions for maintaining the speed or channel capacity of the system under conditions of single component failure.

DESCRIPTION OF RELATED ART INCLUDING INFORMATION DISCLOSED UNDER 37 CFR 1.97 AND 37 CFR 1.98.

[0002] The present invention is a RAID system which has redundant connections between active storage array controllers and the arrays of storage units they control, spare storage units in each array, and a passive storage array controller which assumes the control of the array of storage units of any failed storage array controller. Thus the failure of any one connector, storage unit, or storage array controller does not affect the channel capacity or speed of the RAID system

of this invention.

[0003] U.S. Pat. No. 5,651,110 discloses a RAID system with two second level storage array controllers each of which control an array of disk drives. Each second level storage array controller is controlled by a separate first level storage array controller, which, in turn, communicates with the computer. In the event of a failure of a second level storage array controller, control of the array of disks assigned to the failed storage array controller is assumed by the intact second level storage array controller, which now controls both its original disks and the disks of the failed second level storage array controller. The channel capacity of the RAID system is thereby reduced by half under conditions of a failed second level storage array controller.

[0004] U.S. Pat. No. 5,787,070 discloses a global computer network packet switching system in which a number of active service modules are backed up by a normally passive redundancy module which takes the load when one of the active service modules fails. The communication system has no provisions for data storage.

[0005] U.S. Pat. No. 5,790,775 discloses a data storage system with a SCSI environment. The system involves two storage array controllers in dual-active, redundant configuration, and associated physical storage media. Failure of one storage array controller results in the other storage array controller assuming the control of all of the SCSI units (failover). The reverse operation, wherein the defective storage array controller is repaired or replaced and assumes control of its portion of the storage media, is termed "failback". The channel capacity of the data storage system is reduced by half under conditions of a failed storage array controller.

[0006] U.S. Pat. No. 5,848,230 discloses a RAID system in which there is triple

replication of all subsystems. It has three storage array controllers, one active and two which are normally passive and are used only in case of the failure of the active storage array controller and (subsequently) the secondary storage array controller. In addition, triplicate subsystems such as cooling and power subsystems are included. This system provides highly reliable and continuous availability of storage service and an undiminished channel capacity. The provision of two normally passive storage array controllers for each active storage array controller is a major contributor to the cost of this system.

[0007] U.S. Pat. No. 5,872,906 discloses a RAID system with provisions for allocating a spare disk unit in case of a disk failure. It includes two substorage array controllers which are provided for the common buses thereby distributing the processing functions of the storage array controllers and reducing a load. No provisions for failure of a storage array controller are disclosed.

[0008] U.S. Pat. No. 5,922,077 discloses a RAID system with two storage array controllers and a fail-over switch which routes the data from the storage array controller of a failed communication path to the operating storage array controller, which then handles the load of both storage array controllers. The channel capacity is reduced when one storage array controller is handling both loads.

[0009] U.S. Pat. No. 5,944,838 discloses a RAID system with a redundant storage control module (RDAC) in which two queues of pending I/O requests are maintained for a single array of storage devices. The redundant queue takes over on the failure of the active queue. The redundant queue copies each I/O request sent to the active path which minimizes the time required for the redundant queue to take over the functions of the active queue.

control of the array of storage devices of the failed storage array controller. Since each array of storage units contains a spare unit which becomes active when one storage unit in the array fails, the RAID system of this invention is able to function with undiminished capacity in the event of failure of any one storage unit or any one storage array controller. In another embodiment, each storage unit is connected to controllers by two connectors, and this embodiment is, in addition, able to function in the event of failure of any one connector.

[0013] The objective of this invention is to provide a RAID system with undiminished capacity despite the failure of any one storage unit, any one connector, or any one storage array controller.

[0014] Another objective is to provide a RAID system which produces a signal for the operator in the event of failure of any component.

[0015] Another objective is to provide a RAID system which automatically substitutes a replacement for a failed storage unit or storage array controller.

[0016] Another objective is to provide a RAID system with redundant connectors connecting the storage units and the storage array controllers.

[0017] Another objective is to provide a RAID system with several active storage array controllers which normally control the arrays of storage units and with one passive storage array controller which assumes control the storage units of any active storage array controller which fails.

[0018] Another objective is to provide a RAID system capable of functioning with undiminished channel capacity in the event of failure of a connector, storage unit, or storage array controller with minimal redundancy of components.

[0019] Another objective is to provide a RAID system capable of functioning with undiminished channel capacity in the event of failure of a connector, storage unit, or storage array controller without incurring the expense of a back-up storage array controller for each active storage array controller.

[0020] Another objective is to provide a RAID system capable of functioning with undiminished channel capacity in the event of failure of a connector, storage unit, or storage array controller at minimal expense.

[0021] A final objective is to produce a RAID system simply constructed of inexpensive, readily obtainable components without adverse effects on the environment.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS.

[0022] Fig. 1 is a diagrammatic depiction of a single RAID subsystem.

[0023] Fig. 2 is a diagrammatic depiction of a redundant single RAID subsystem.

[0024] Fig. 3 is a diagrammatic depiction of the first embodiment RAID system of this invention having three active storage array controllers and one passive storage array controller.

[0025] Fig. 4 is a diagrammatic depiction of the second embodiment RAID system of this invention having three active storage array controllers and one passive storage array controller.

[0026] Fig. 5 is a diagrammatic depiction of the third embodiment RAID system of this invention having n active storage array controllers and one passive storage array controller.

[0027] Fig. 6 is a diagrammatic depiction of the fourth embodiment RAID system of this invention having n active storage array controllers and one passive storage array controller.

[0028] Fig. 7 is a diagrammatic depiction of the fifth embodiment RAID system of this invention having two active storage array controllers and one passive storage array controller.

[0029] Fig. 8 is a flow chart showing the process of detecting the failure of an active storage array controller by an adjacent storage array controller, assuming the duties of the failed active storage array controller by the passive storage array controller, and signaling the occurrence of a failure.

[0030] Fig. 9 is a flow chart showing the process of detecting the failure of an active storage array controller by the passive storage array controller, assuming the duties of the failed active storage array controller by the passive storage array controller, and signaling the occurrence of a failure

DETAILED DESCRIPTION OF THE INVENTION.

[0031] In this patent application the term "channel capacity" means the ability of a given channel subject to specific constraints to transmit messages from a specified message source expressed as the maximum possible average transinformation rate, which can be achieved with an arbitrary small probability of errors by use of an appropriate code. The channel capacity of a RAID system is commonly referred to as the "speed" of the system.

[0032] Fig. 1 is a schematic of the external view of a RAID system referred to in this application as a "single RAID subsystem" 11. The single RAID subsystem comprises a storage array controller 30, and an array of direct access storage devices (DASD) or storage units 40-61. A host computer is electrically connected to the storage array controller 30 by connector 20.

[0033] Any suitable connector may be used, such as a wire, copper wire, cable, optical fiber, or a SCSI bus.

[0034] In all of the Figs. the convention is followed of depicting connectors which are not electrically connected as lines which cross perpendicularly. An electrical connection is indicated

by a line which terminates perpendicularly at another line or at a symbol for a component. Thus in Fig. 1 a host computer (not shown in Fig. 1) is electrically connected to storage array controller 30 by connector 20. The host computer is not considered part of the single RAID subsystem and is not shown in Fig. 1. Connector 401 is electrically connected to storage array controller 30 and to DASD 1A 40 and to DASD 1B 41, but is not electrically connected to connectors 402 to 406. Connector 402 connects storage array controller 30 with DASD 2A 42 and DASD 2B 43. Connector 403 connects storage array controller 30 with DASD 3A 44 and DASD 3B 45. Connector 404 connects storage array controller 30 with DASD 4A 46 and DASD 4B 47. Connector 405 connects storage array controller 30 with DASD 5A 50 and DASD 5B 51. Connector 406 connects storage array controller 30 with DASD 6A 60 and DASD 6B 61.

[0035] In the configuration in Fig. 1, for example, data are striped over DASD 1A 40, DASD 2A 42, DASD 3A 44, and DASD 4A 46. DASD 5A 50 is a parity disk which is used to check the accuracy of data striped on the disks DASD 1A-4A and to substitute for a failed DASD. DASD 6A is a spare disk which is used to substitute for any one of disks DASD 1A-5A which have failed.

[0036] DASD may be disks, tapes, CDS, or other suitable storage device. A preferred DASD is a disk.

[0037] All the storage units or DASD and connectors in a system taken as a whole is referred to as an "array" of storage units or DASD, respectively. In the examples here the DASD are arranged in channels which consist of a number of DASD which are electrically connected to each other and to the storage array controller by connectors. The channels are designated in Fig. 1 as 1-6. The number of channels may vary. A preferred number of channels is 6.

[0038] A channel, for example channel 1, consists of connector 401, DASD 1A 40, and DASD 1B 41. Although only two DASD are depicted in channel 1 of Fig. 1, there may be as many as 126 DASD in a channel. A preferred number of DASD in a channel is five.

[0039] A group of DASDs served by separate channels across which data is striped is referred to as a "tier" of DASDs. A DASD may be uniquely identified by a channel number and a tier letter, for example DASD 1A is the first disk connected to channel 1 and tier A of the storage array controller.

[0040] A preferred storage array controller is the Fibre Sabre 2100 Fibre Channel RAID storage array controller manufactured by Digi-Data Corporation, Jessup, Maryland.

[0041] Additional tiers of DASDs may be used.

[0042] Any suitable host computer may be used. A preferred host computer is a PENTIUM microchip-based personal computer available from multiple vendors such as IBM, Research Triangle park, North Carolina; Compaq Computer Corp., Houston Texas; or Dell Computer, Austin, Texas. PENTIUM is a trademark for microchips manufactured by Intel Corporation, Austin, Texas.

[0043] Fig. 2 is a schematic of a RAID system referred to in this application as a "redundant single RAID subsystem" 21. The redundant single RAID subsystem 21 is identical to the single RAID subsystem 11 of Fig. 1 except that each DASD in the redundant single RAID subsystem is connected to the storage array controller 30 by two connectors. Connector 501 is connected to disk array storage array controller 30 and to DASD 1A 40 and DASD 1B 41. Connector 502 connects storage array controller 30 with DASD 2A 42 and DASD 2B 43. Connector 503 connects storage array controller 30 with DASD 3A 44 and DASD 3B 45.

Connector 504 connects storage array controller 30 with DASD 4A 46 and DASD 4B 47.

Connector 505 connects storage array controller 30 with DASD 5A 50 and DASD 5B 51.

Connector 506 connects storage array controller 30 with DASD 6A 60 and DASD 6B 61.

[0044] The single RAID subsystem 11 in Fig. 1, and redundant single RAID subsystem 21 in Fig. 2 therefore are protected against failure of any two disks, by the inclusion of a parity disk DASD 5A 50 and DASD 5B 51 and by the inclusion of a spare disk DASD 6A 60 and DASD 6B 61 in each channel. The redundant single RAID subsystem 21 in Fig. 2 is protected against failure of any single connector which connects a DASD to the storage array controller 30 by the inclusion of two connectors, for example 401 and 501, which connect each DASD, for example DASD 1A 40, to the storage array controller 30.

[0045] Fig. 3 shows the first embodiment RAID system of the present invention. In this system, the storage array controller of redundant RAID subsystem 21 is connected to the storage array controller of redundant RAID subsystem 121 by two connectors, depicted in Fig. 3 as connectors 114 and 116. All connectors in Fig. 3 are bidirectional connectors. Subsystem 121 is connected to subsystem 221 by connectors 118 and 120. Subsystem 221 is connected to subsystem 21 by connectors 122 and 124. Subsystems 21, 121, and 221 each has an array of DASD and are used for normal RAID functions. Each storage array controller of subsystems 21, 121, and 221 therefore is attached to two adjacent storage array controllers, forming a ring of storage array controllers. The storage array controllers for subsystems 21, 121, and 221 are referred to as "active" storage array controllers because in the normal function of the RAID system these storage array controllers are actively involved in controlling the arrays of DASD in reading and writing data.

[0046] Storage array controller 100 is similar to the storage array controllers of subsystems 21, 121, and 221 except that it is not normally associated with an array of DASD. Storage array controller 100 is a "passive" storage array controller and serves as a back-up for the storage array controllers associated with subsystems 21, 121, and 221. Storage array controller 100 is connected to the storage array controller of subsystem 21 by connectors 102 and 104; to the storage array controller of subsystem 121 by connectors 110 and 112; and to the storage array controller of subsystem 221 by connectors 106 and 108.

[0047] The storage array controller of subsystems 21, 121, and 221 contain internal software which generates a binary signal termed a "normal operating signal" or a "heartbeat" at an interval of a few milliseconds when the storage array controllers of the respective subsystems are operational. When the storage array controller is in a defective condition, the emission of the normal operating signal ceases.

[0048] The normal operating signal is emitted from the storage array controller of subsystem 21 over connector 114 to the disk array storage array controller of subsystem 121. In similar fashion, the normal operating signal is emitted from the storage array controller of subsystem 121 over connector 118 to the storage array controller of subsytem 221. Finally, the normal operating signal is emitted from the storage array controller of subsystem 221 over connector 122 to the storage array controller of subsystem 21.

[0049] When one storage array controller, for example 121, no longer receives the normal operating signal because the storage array controller of the adjacent subsystem, 21 in this example, is defective, the receiving storage array controller 121, referred to as the "reporter" storage array controller, emits an activation signal to the passive storage array controller 100. On

receipt of this activation signal, the passive storage array controller 100 assumes the identify of the failed storage array controller of subsystem 21. The passive storage array controller consults a table stored on each DASD and identifies the DASD of the defective storage array controller 21. The passive storage array controller 100 then assumes control of the DASD array of subsystem 21.

[0050] In the first embodiment depicted in Fig. 3 each active controller 21, 121, and 221 also emits a heartbeat to the passive controller 100 over connectors 102, 110, and 106, respectively. Failure of the heartbeat from any one active controller also activates the passive controller to assume identify of the failed controller, as described above. This system provides redundancy in that the passive controller is signaled concerning the failure of an active controller by both indirectly by the reporter storage array controller and directly by the failure of the heartbeat from the defective active storage controller.

[0051] Finally, failure of an active storage array controller 21 causes the reporting storage array controller 121 or the passive storage array controller 100 to emit a warning signal which indicates to the operator of the RAID system that a storage array controller has failed and requires repair or replacement.

[0052] Although a single connector is described in the above example, each disk array storage array controller may be connected to adjacent storage array controllers by two redundant connectors. This assures that the failure of one connector between storage array controllers will not result in the loss of communication between the storage array controllers, because the other connector will convey the signal. In a similar fashion, the passive storage array controller 100 may be connected to the storage array controllers of subsystems 21, 121, and 221 by two

redundant connectors which assure communications in the event of failure of any one connector. Thus the passive storage array controller is able to assume the identify and function of any of the active storage array controllers even in the event of the failure of any one connector between the passive storage array controllers and the active storage array controllers.

[0053] Fig. 4 shows a second embodiment of the RAID system of the present invention. The second embodiment shown in Fig. 4 is the same as the first embodiment shown in Fig. 3 except that the connectors 114-124 are omitted. The second embodiment has the advantage of lesser costs, as compared to the first embodiment, but, on the other hand, the second embodiment lacks the redundance afforded by the first embodiment.

[0054] Fig. 5 depicts a third embodiment RAID system with n redundant single RAID subsystems. The system of Fig. 5 is the same as that of Fig. 3 except for the inclusion of additional redundant single RAID subsystems represented by redundant single RAID subsystems n-1 and n. In the embodiment of Fig. 5, the connections between redundant single RAID subsystems 21, 121, and 221 with passive storage array controller 100 are as in Fig. 3 with the exceptions that subsystem 121 is connected to subsystem n-1 by connectors 130 and 132, and subsystem 221 is connected to subsystem n by connectors 146 and 148. Subsystem n-1 is connected to subsystem n by connectors 138 and 140. Passive storage array controller 100 is connected to subsystem n-1 by connectors 134 and 136. Passive storage array controller 100 is connected to subsystem n by connectors 142 and 144.

[0055] Fig. 6 shows a fourth embodiment of the RAID system of the present invention. The fourth embodiment shown in Fig. 6 is the same as the third embodiment shown in Fig. 5 except that the connectors 114, 116, 122, 124, 130, 132, 138, 140, 146, 148, are omitted. The

fourth embodiment has the advantage of lesser costs, as compared to the third embodiment, but, on the other hand, the fourth embodiment lacks the redundance afforded by the third embodiment.

[0056] In the first to fourth embodiments of the RAID system described above and shown in Figs. 3-6, a component of each embodiment is the redundant single RAID subsystem, 21 in Fig. 2. The single RAID subsystem, 11 in Fig. 1, may be substituted for the redundant single RAID subsystem in the first to fourth embodiments. It should be noted that the failure of the single connector which connects the DASD of an array in a single RAID subsystem, 11 in Fig. 1, as incorporated in embodiments two to four of the RAID system, does not affect the channel capacity of the RAID systems of this invention. Failure of the single connector in a single RAID subsystem, 11 in Fig. 1, does affect the ability to access and store data in the DASD of the array. The loss of data access on the failure of a single connector is avoided in the redundant single RAID subsystem, 21 in Fig. 2, which has two connectors between each DASD and storage array controller.

[0057] Fig. 7 shows a fifth embodiment of the RAID system of the present invention. In Fig. 7, there are two active storage array controllers, 620 and 640, and one passive storage array controller 600. Active storage array controller 620 is connected to passive storage array controller 600 by connectors 606 and 608. Active storage array controller 640 is connected to passive storage array controller 600 by connectors 607 and 609.

[0058] Active storage array controller 620 has in its first channel dual-ported DASD 621-625 and is connected to these DASD by connector 631. Active storage array controller 620 has in its second channel dual-ported DASD 626-630 and is connected to these DASD by connector 632.

[0059] Active storage array controller 640 has in its first channel dual-ported DASD 646-650 and is connected to these DASD by connector 641. Active storage array controller 640 has in its second channel dual-ported DASD 651-655 and is connected to these DASD by connector 642.

[0060] Connector 631 is also connected to the dual-ported DASD 646-650. Connector 632 is also connected to the dual-ported DASD 651-655.

[0061] Connector 641 is also connected to the dual-ported DASD 621-625. Connector 642 is also connected to the dual-ported DASD 626-630.

[0062] Passive storage array controller 600 is connected by connector 602 to connector 631 and by connector 604 to connector 632.

[0063] In the fifth embodiment RAID system, therefore, each of the two active storage array controllers, 620 and 640, and the passive storage array controller 600, are connected to each of the DASD. Each DASD has two connectors leading directly or indirectly to the active and passive controllers. Failure of either of the active controllers or one of the connectors leading to the DASD will result in the assumption of control of the DASD involved by the passive storage array controller.

[0064] Fig. 8 is a diagram of the process which occurs after failure of a storage array controller described above with reference to the first embodiment depicted in Fig. 3. The failure of the defective storage array controller 121 of a redundant single RAID subsystem halts the emission of the heartbeat in step 510. The adjacent storage array controller 221, termed the "reporter" storage array controller, notes the cessation of the heartbeat emitted by the defective storage array controller and emits an activation signal 520. Using its associated interface chip, the

passive storage array controller 100 assumes the identity of the defective storage array controller 121 in step 530. The passive storage array controller 100 also identifies the DASD of the defective storage array by reading a table of DASD addresses and storage array controller assignments previously stored on each DASD 540. The newly activated passive storage array controller 100 assumes control of the DASD of the defective storage array controller 121 in step 550. Finally, the reporter storage array controller 221 or the newly activated passive storage array controller 100 emits a warning signal to alert the operator of the RAID system to the need for repair or replacement of the defective storage array controller in step 560.

[0065] Fig. 9 is a diagram of an alternate process which occurs after failure of a storage array controller described above with reference to the first embodiment depicted in Fig. 3. The process is the same as in Fig. 8 except that step 520 is deleted and in step 580 the passive storage array controller 100 detects the cessation of the heart beat. In all other respects the process in Fig. 9 is identical to that in Fig. 8.

[0066] Although the above example depicts a RAID system having three redundant single RAID subsystem and one passive storage array controller, the number of redundant single RAID subsystems may range from 1 to n, where n is a arbitrary number, as in Figs.5 and 6. A preferred range for n is 2-20. Of course, the larger n is, the lower the relative additional cost of including the passive storage array controller in the system. On the other hand, the larger n is, the greater is the chance, however remote, that greater than one storage array controller will fail before the first failed storage array controller is repaired or replaced by the operator.

[0067] The RAID system of this invention is characterized by undiminished channel capacity despite failure of a DASD, connector, or storage array controller. Thus the channel

capacity (C) of a RAID system comprised of n single RAID subsystems, or n redundant single RAID subsystems, each of which has the same capacity c, is C=(n)(c) despite the failure of any one of a DASD, connector or storage array controller. This is an important advantage over conventional RAID systems because the conventional RAID systems suffer diminished capacity when a connector, or storage array controller fails. In particular, if a conventional RAID system comprises two subsystems and operates without a failed component at a capacity C=2(c) and one storage array controller or connector fails, the capacity of the RAID system becomes C=c. Thus in this example the capacity is reduced by half by the failure of a connector or storage array controller.

[0068] In the more general case, if a conventional RAID system comprises n subsystems, the capacity of the normally operating system is C=(n)(c), and the capacity after the failure of a subsystem is C=(n-1)(c). Thus the capacity is reduced by a factor related to the number of subunits by the failure of a subsystem.

[0069] It will be apparent to those skilled in the art that the examples and embodiments described herein are by way of illustration and not of limitation, and that other examples may be used without departing from the spirit and scope of the present invention, as set forth in the appended claims.